

Matthew Thurston  
Mini-project proposal  
9/11/2025

**Research Question:** Do collective investor sentiment on Reddit, StockTwits, and Google search trends do a good job in predicting next-day stock-market returns or volatility?

**Abstract:**

With retail investors increasingly shaping market swings, I plan to examine whether crowd psychology visible on social platforms translates into measurable market movement. By collecting daily sentiment and attention data from Reddit (r/wallstreetbets, r/stocks), StockTwits messages, and Google Trends search interest for a selected basket of U.S. stocks, I can test whether shifts in online mood precede changes in market prices or volatility. This project will bridge data science and psychology, providing insight into how herding behavior and emotional contagion influence trading decisions.

**Data:**

The data I plan to use is highly tractable and mostly public:

- Reddit posts and comments via the official API (PRAW) for specified subreddits.
- StockTwits messages by ticker through its public REST API.
- Google Trends daily search-interest scores for company names and ticker symbols using the pytrends library.
- Market Data (daily prices, returns, and volatility) from Yahoo Finance or Alpha Vantage.

All three sentiment sources can be collected programmatically and aligned to market close times for analysis.

**Strategy:**

I will create a daily dataset merging sentiment, message volume, and Google search interest with market outcomes.

Steps:

- Scrape & preprocess text to extract ticker mentions and clean for language and bots.
- Sentiment scoring with VADER or FinBERT to quantify bullish vs. bearish tone.
- Feature engineering: daily average sentiment, message-volume z-scores, and Google Trends anomalies per ticker.
- Modeling: time-series regression and classification (e.g., logistic regression, random forest) to predict next-day returns or volatility, using walk-forward cross-validation to avoid look-ahead bias.

This approach will let me test if sentiment shifts Granger-cause market movements.

**Stakeholders:**

- Retail investors and professional traders can gain insight into how collective mood shapes prices.
- Financial analysts and risk managers could use findings to improve forecasting models.
- Individual Investors could use these findings to invest with data-driven decisions
- Platform operators (Reddit, StockTwits) and market regulators (SEC, FINRA) may also be interested in understanding how social sentiment drives trading activity.

**Ethical Concerns:**

I will follow each platform's Terms of Service and API rate limits, avoid collecting personally identifiable information, and store only aggregated sentiment metrics.

The project is for academic purposes and will emphasize transparency and reproducibility while respecting privacy and data-usage policies.